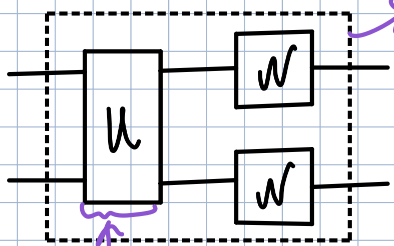
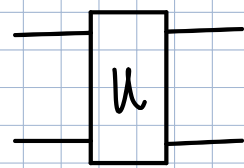
# ① Recap: Quantifying noise

$\tilde{\mathcal{U}}(\cdot) = (\mathcal{N} \otimes \mathcal{N}) \circ \mathcal{U}.$ $\qquad\qquad$ $\mathcal{U}(\cdot) = U(\cdot)U^\dagger$

Noisy gate




VS.




ideal gate. $\qquad$ $\mathcal{N}(\rho) = (1-\underline{\underline{p}})\rho + p\,\mathrm{Tr}[\rho]\frac{\mathbb{1}}{2}$ $\qquad$ ideal gate.

Quantify the error by the diamond norm distance b/w the ideal and noisy gate
$\underline{\underline{\frac{1}{2}\|\tilde{\mathcal{U}} - \mathcal{U}\|_\diamond}}$.

→ Operational meaning: optimal worst-case error over all possible "tests" of the ideal and noisy gates.

Trace norm — contains optimization over all measurements.

$$\max_{|\psi\rangle} \frac{1}{2}\left\| (id \otimes \tilde{\mathcal{U}})(|\psi\rangle\langle\psi|) - (id \otimes \mathcal{U})(|\psi\rangle\langle\psi|) \right\|_1 =: \underline{\underline{\frac{1}{2}\|\tilde{\mathcal{U}} - \mathcal{U}\|_\diamond}}$$

input state

Diamond Norm (norm on superoperators)

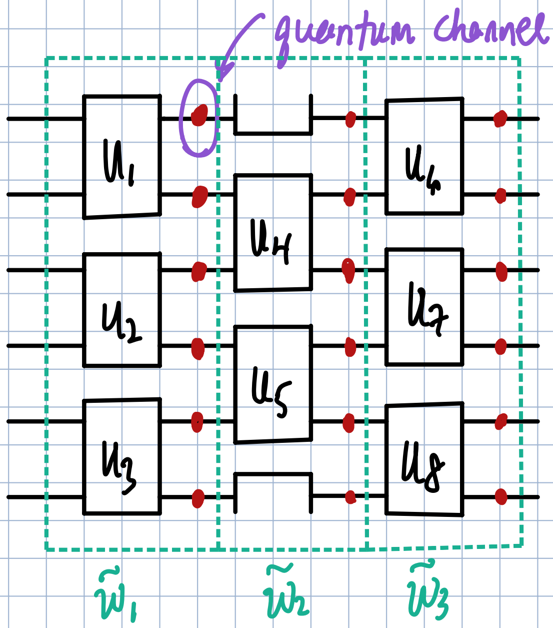
(Test: combination of input state and measurement.)

→ This depends on the ideal gate. For a gate independent quantity, recall from last time that $\boxed{\frac{1}{2}\|\tilde{\mathcal{U}} - \mathcal{U}\|_\diamond \le \|\mathcal{N} - id\|_\diamond.}$

$P\mathcal{N}(P\rho P)P = \mathcal{N}(\rho) \quad \forall P \in$ Pauli

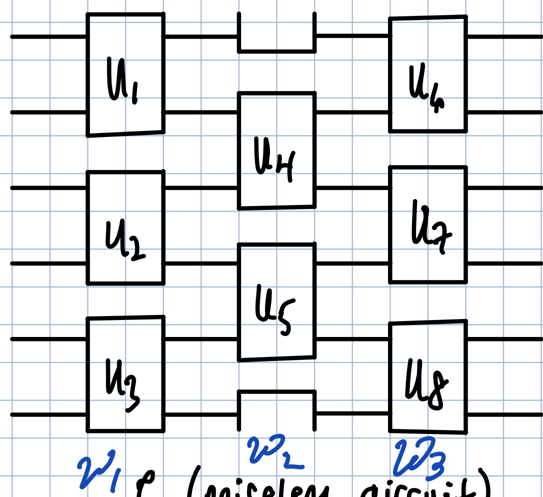
- Depolarizing: $\mathcal{N}(\rho) = (1-p)\rho + \frac{p}{3}(X\rho X + Y\rho Y + Z\rho Z) = q\rho + (1-q)\mathrm{Tr}[\rho]\frac{\mathbb{1}}{2} \Rightarrow \underline{\underline{\frac{1}{2}\|\mathcal{N} - id\|_\diamond = p}}.$
- Dephasing: $\mathcal{N}(\rho) = (1-p)\rho + pZ\rho Z \Rightarrow \underline{\underline{\frac{1}{2}\|\mathcal{N} - id\|_\diamond = p}}.$
- Amplitude damping: $\mathcal{N}(\rho) = K_1\rho K_1^\dagger + K_2\rho K_2^\dagger$, $K_1 = \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{1-\gamma} \end{pmatrix}$, $K_2 = \begin{pmatrix} 0 & \sqrt{\gamma} \\ 0 & 0 \end{pmatrix}$
$\underline{\underline{\frac{1}{2}\|\mathcal{N} - id\|_\diamond = \gamma}}.$

# ② Accumulation of Errors.




$\tilde{\mathcal{E}}$ (noisy circuit)

VS.




$\mathcal{E}$ (noiseless circuit).

$$\tilde{\mathcal{E}} = \tilde{\mathcal{W}}_3 \circ \tilde{\mathcal{W}}_2 \circ \tilde{\mathcal{W}}_1 , \quad \mathcal{E} = \mathcal{W}_3 \circ \mathcal{W}_2 \circ \mathcal{W}_1$$

layers.

$$\frac{1}{2}\|\tilde{\mathcal{E}} - \mathcal{E}\|_\diamond = \frac{1}{2}\|\tilde{\mathcal{W}}_3 \circ \tilde{\mathcal{W}}_2 \circ \tilde{\mathcal{W}}_1 - \mathcal{W}_3 \circ \mathcal{W}_2 \circ \mathcal{W}_1\|_\diamond$$

$$= \frac{1}{2}\|\tilde{\mathcal{W}}_3 \circ \tilde{\mathcal{W}}_2 \circ \tilde{\mathcal{W}}_1 - \mathcal{W}_3 \circ \tilde{\mathcal{W}}_2 \circ \tilde{\mathcal{W}}_1 + \mathcal{W}_3 \circ \tilde{\mathcal{W}}_2 \circ \tilde{\mathcal{W}}_1 - \mathcal{W}_3 \circ \mathcal{W}_2 \circ \mathcal{W}_1\|_\diamond$$

$\|\mathcal{N}_1 \circ \mathcal{N}_2\|_\diamond \le \|\mathcal{N}_1\|_\diamond \|\mathcal{N}_2\|_\diamond$

$$\le \frac{1}{2}\|(\tilde{\mathcal{W}}_3 - \mathcal{W}_3) \circ \tilde{\mathcal{W}}_2 \circ \tilde{\mathcal{W}}_1\|_\diamond + \frac{1}{2}\|\mathcal{W}_3 \circ (\tilde{\mathcal{W}}_2 \circ \tilde{\mathcal{W}}_1 - \mathcal{W}_2 \circ \mathcal{W}_1)\|_\diamond$$

$$\le \frac{1}{2}\|\tilde{\mathcal{W}}_3 - \mathcal{W}_3\|_\diamond \underbrace{\|\tilde{\mathcal{W}}_2 \circ \tilde{\mathcal{W}}_1\|_\diamond}_{=1} + \frac{1}{2}\underbrace{\|\mathcal{W}_3\|_\diamond}_{=1} \|\tilde{\mathcal{W}}_2 \circ \tilde{\mathcal{W}}_1 - \mathcal{W}_2 \circ \mathcal{W}_1\|_\diamond$$

$$= \frac{1}{2}\|\tilde{\mathcal{W}}_3 - \mathcal{W}_3\|_\diamond + \frac{1}{2}\underbrace{\|\tilde{\mathcal{W}}_2 \circ \tilde{\mathcal{W}}_1 - \mathcal{W}_2 \circ \mathcal{W}_1\|_\diamond}$$

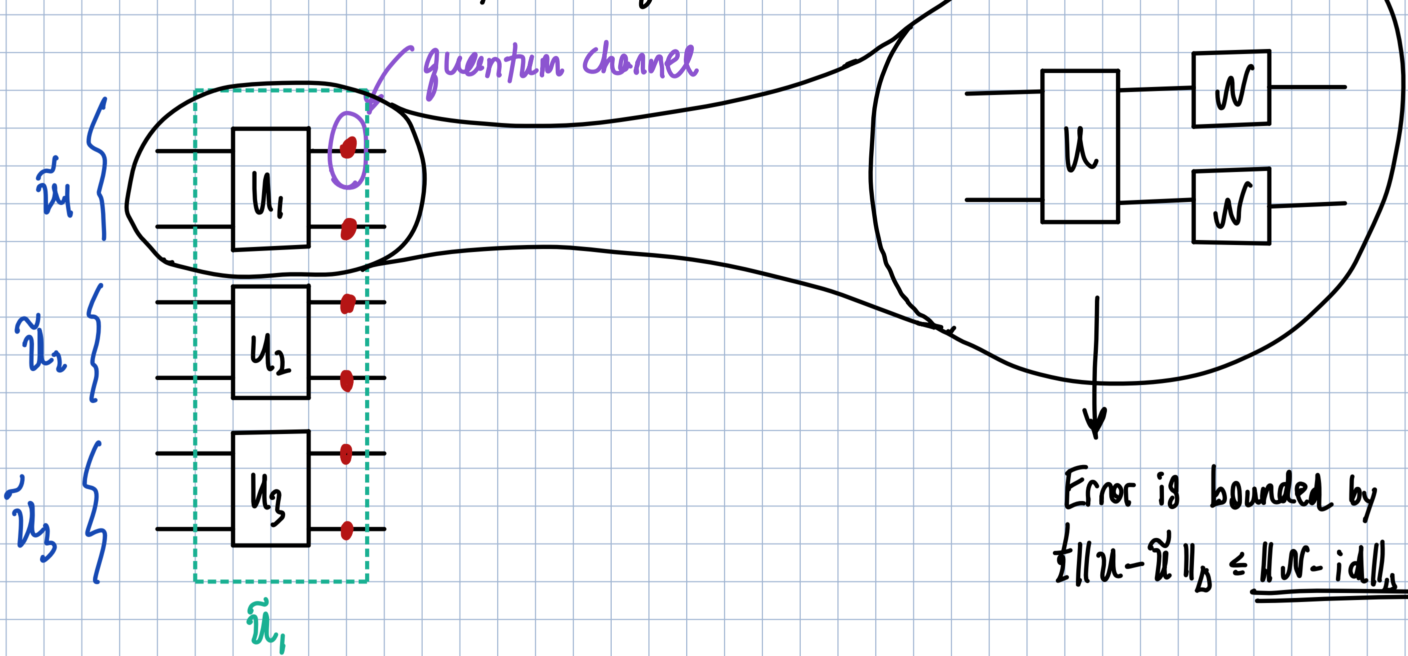
$$= \|\tilde{\mathcal{W}}_2 \circ \tilde{\mathcal{W}}_1 - \mathcal{W}_2 \circ \tilde{\mathcal{W}}_1 + \mathcal{W}_2 \circ \tilde{\mathcal{W}}_1 - \mathcal{W}_2 \circ \mathcal{W}_1\|_\diamond$$

$$\le \|\tilde{\mathcal{W}}_2 - \mathcal{W}_2\|_\diamond + \|\tilde{\mathcal{W}}_1 - \mathcal{W}_1\|_\diamond$$

$$\le \frac{1}{2}\|\tilde{\mathcal{W}}_3 - \mathcal{W}_3\|_\diamond + \frac{1}{2}\|\tilde{\mathcal{W}}_2 - \mathcal{W}_2\|_\diamond + \frac{1}{2}\|\tilde{\mathcal{W}}_1 - \mathcal{W}_1\|_\diamond.$$

Errors add up over the layers!

Now consider one of the layers: $\frac{1}{2}\|\tilde{W}_1 - W_1\|_\diamond$.




Error is bounded by
$\frac{1}{2}\|\mathcal{U} - \tilde{\mathcal{U}}\|_\diamond \le \underline{\underline{\|\mathcal{N} - \mathrm{id}\|_\diamond}} \equiv \kappa$.

$$\tilde{W}_1 = \tilde{\mathcal{U}}_1 \otimes \tilde{\mathcal{U}}_2 \otimes \tilde{\mathcal{U}}_3$$

$$\frac{1}{2}\|\tilde{W}_1 - W_1\|_\diamond = \frac{1}{2}\|\tilde{\mathcal{U}}_1 \otimes \tilde{\mathcal{U}}_2 \otimes \tilde{\mathcal{U}}_3 - \mathcal{U}_1 \otimes \mathcal{U}_2 \otimes \mathcal{U}_3\|_\diamond$$

$$= \frac{1}{2}\|\tilde{\mathcal{U}}_1 \otimes \tilde{\mathcal{U}}_2 \otimes \tilde{\mathcal{U}}_3 - \mathcal{U}_1 \otimes \tilde{\mathcal{U}}_2 \otimes \tilde{\mathcal{U}}_3 + \mathcal{U}_1 \otimes \tilde{\mathcal{U}}_2 \otimes \tilde{\mathcal{U}}_3 - \mathcal{U}_1 \otimes \mathcal{U}_2 \otimes \mathcal{U}_3\|_\diamond$$

$\|\mathcal{N}_1 \otimes \mathcal{N}_2\|_\diamond = \|\mathcal{N}_1\|_\diamond \cdot \|\mathcal{N}_2\|_\diamond$

$$\le \frac{1}{2}\|(\tilde{\mathcal{U}}_1 - \mathcal{U}_1) \otimes \tilde{\mathcal{U}}_2 \otimes \tilde{\mathcal{U}}_3\|_\diamond + \frac{1}{2}\|\mathcal{U}_1 \otimes (\tilde{\mathcal{U}}_2 \otimes \tilde{\mathcal{U}}_3 - \mathcal{U}_2 \otimes \mathcal{U}_3)\|_\diamond$$

$$= \underbrace{\|\tilde{\mathcal{U}}_1 - \mathcal{U}_1\|_\diamond}_{\le 2\kappa} \underbrace{\|\tilde{\mathcal{U}}_2\|_\diamond}_{=1} \underbrace{\|\tilde{\mathcal{U}}_3\|_\diamond}_{=1}$$

$$= \underbrace{\|\mathcal{U}_1\|_\diamond}_{=1} \|\tilde{\mathcal{U}}_2 \otimes \tilde{\mathcal{U}}_3 - \mathcal{U}_2 \otimes \mathcal{U}_3\|_\diamond$$

$$= \|\tilde{\mathcal{U}}_2 \otimes \tilde{\mathcal{U}}_3 - \mathcal{U}_2 \otimes \tilde{\mathcal{U}}_3 + \mathcal{U}_2 \otimes \tilde{\mathcal{U}}_3 - \mathcal{U}_2 \otimes \mathcal{U}_3\|_\diamond$$

$$\le \underbrace{\|\tilde{\mathcal{U}}_2 - \mathcal{U}_2\|_\diamond}_{\le 2\kappa} \|\tilde{\mathcal{U}}_3\|_\diamond + \|\mathcal{U}_2\|_\diamond \underbrace{\|\tilde{\mathcal{U}}_3 - \mathcal{U}_3\|_\diamond}_{\le 2\kappa}$$

$\le 3\kappa \rightarrow \left(\frac{n}{2}\right)\kappa \le n\kappa$

($n \equiv$ # of qubits).

$$\Rightarrow \boxed{\frac{1}{2}\|\tilde{\mathcal{E}} - \mathcal{E}\|_\diamond \le L n \kappa}$$

$L$ → # of Layers,

$n$ → # of qubits.

$\kappa$ → noise rate.

- <u>Observe</u>: $n \cdot L$ is simply the number of noise channels in the circuit!

- Fix $n$, $\kappa$. Fix a desired error $\varepsilon \in [0,1)$. What is the maximum possible depth?

$$\frac{1}{2}\|\tilde{\mathcal{E}}-\mathcal{E}\|_\diamond \lesssim Ln\kappa \le \varepsilon \Rightarrow L \le \frac{\varepsilon}{n\kappa}$$

inversely proportional to noise rate! (higher noise ⇒ lower depth.)

- Now fix $L$. What is the maximum tolerable noise rate to achieve error $\varepsilon$?

$$Ln\kappa \le \varepsilon \Rightarrow \kappa \le \frac{\varepsilon}{Ln}$$

(higher depth ⇒ lower tolerable noise).

- Now consider depth as a function of # of qubits.

(*) <u>Shallow circuits</u>: Constant-depth $\rightarrow L = O(1)$

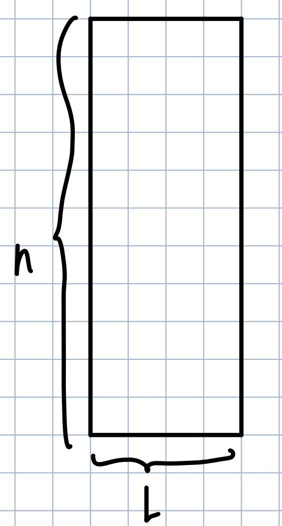
(Sometimes: log-depth $\rightarrow L = O(\log n)$.)

(Depth grows slower than number of qubits — or is independent of # of qubits).

**<u>Aside</u>: Big-O notation.**

$f(x) = O(g(x))$

$\rightarrow |f(x)| \le \alpha |g(x)|$, $\alpha > 0$, $x \to \infty$




Science

Current Issue

HOME > SCIENCE > VOL. 362, NO. 6412 > QUANTUM ADVANTAGE WITH SHALLOW CIRCUITS

REPORT

**Quantum advantage with shallow circuits**

SERGEY BRAVYI, DAVID GOSSET, AND ROBERT KÖNIG   Authors Info & Affiliations

*SCIENCE* • 19 Oct 2018 • Vol 362, Issue 6412 • pp. 308-311 • DOI: 10.1126/science.aar3106

nature physics

Explore content ⌄   About the journal ⌄   Publish with us ⌄

nature > nature physics > articles > article

Article | Published: 06 July 2020

**Quantum advantage with noisy shallow circuits**

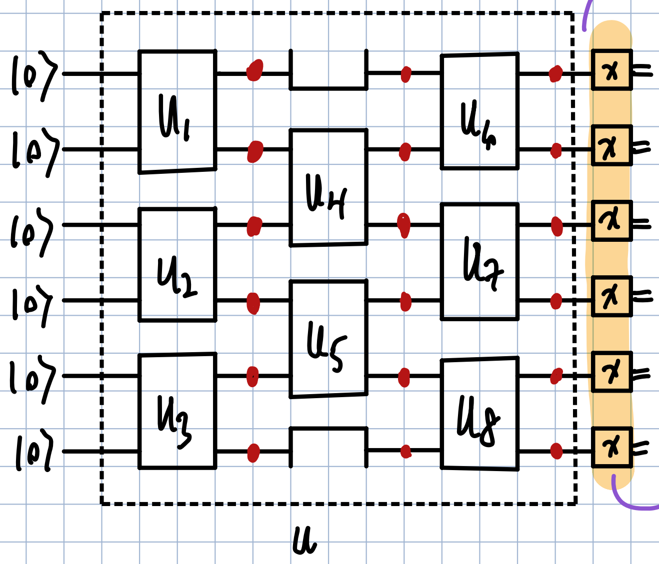
Sergey Bravyi, David Gosset, Robert König & Marco Tomamichel

*Nature Physics* **16**, 1040–1045 (2020) | Cite this article

# ③ Diving Deeper

- Consider specific classes of circuits (e.g., shallow, random).
- Consider specific problems. → e.g., Pauli expectation values

State just before measurement: $\rho$




$P \in \mathcal{P}_n = \{\mathbb{1}, X, Y, Z\}^{\otimes n}$

$\mathrm{Tr}[\rho P]$ $\quad P = X \otimes Y \otimes Z \otimes Y$

Measure in some n-qubit Pauli basis.

This task is quite general!

1. Average value of energy : $\mathrm{Tr}[H \rho]$ — Hamiltonian.

Decompose $H$ as $H = \sum_{P \in \mathcal{P}_n} c_P P$, $\quad c_P = \frac{1}{2^n} \mathrm{Tr}[PH]$. → We can always do this, b/c the Paulis form a basis!

$\mathcal{P}_n = \{\mathbb{1}, X, Y, Z\}^{\otimes n} \to 4^n$

$$X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad Y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

Then $\mathrm{Tr}[H\rho] \longrightarrow = \sum_{P \in \mathcal{P}_n} c_P \, \mathrm{Tr}[P\rho]$ → Pauli expectation value!

↳ $H$ is given in most applications via these coefficients.

2. <u>Computational basis measurement</u>: $\underline{\mathrm{Tr}[|\vec{x}\rangle\langle\vec{x}|\rho]}$ (meas. outcome probabilities).
$\vec{x} \in \{0,1\}^n$

<u>Lemma</u>: $|\vec{x}\rangle\langle\vec{x}| = \frac{1}{2^n}\sum_{\vec{y}\in\{0,1\}^n} (-1)^{\vec{x}\cdot\vec{y}} Z^{\vec{y}}$   $\vec{x} = (0,1,1,0,1)$

$\vec{y} = (y_1, y_2, \ldots, y_n) \Rightarrow Z^{\vec{y}} = Z^{y_1} \otimes Z^{y_2} \otimes \cdots \otimes Z^{y_n}$

$\vec{x}\cdot\vec{y} = (x_1,\ldots,x_n)\cdot(y_1,\ldots,y_n)$   $Z^0 = \mathbb{1},\ Z^1 = Z.$

$= x_1y_1 + \cdots + x_ny_n$

<u>Proof</u>: $Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} = |0\rangle\langle 0| - |1\rangle\langle 1| \Rightarrow |0\rangle\langle 0| = \frac{\mathbb{1}+Z}{2},\ |1\rangle\langle 1| = \frac{\mathbb{1}-Z}{2}.$

$= \frac{1}{2}\left(\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}\right) = \frac{1}{2}\begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} = |0\rangle\langle 0|$ ✓

$= \frac{1}{2}\left(\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}\right) = \frac{1}{2}\begin{pmatrix} 0 & 0 \\ 0 & 2 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} = |1\rangle\langle 1|$ ✓

$\Rightarrow x \in \{0,1\} \Rightarrow |x\rangle\langle x| = \frac{1}{2}\left(\mathbb{1} + (-1)^x Z\right) = \frac{1}{2}\sum_{y\in\{0,1\}} (-1)^{xy} Z^y$

$\vec{x} \in \{0,1\}^n \Rightarrow |\vec{x}\rangle\langle\vec{x}| = |x_1\rangle\langle x_1| \otimes |x_2\rangle\langle x_2| \otimes \cdots \otimes |x_n\rangle\langle x_n|$

$= \left(\frac{1}{2}\sum_{y_1\in\{0,1\}} (-1)^{x_1y_1} Z^{y_1}\right) \otimes \left(\frac{1}{2}\sum_{y_2\in\{0,1\}} (-1)^{x_2y_2} Z^{y_2}\right) \otimes \cdots$

$\otimes \left(\frac{1}{2}\sum_{y_n\in\{0,1\}} (-1)^{x_ny_n} Z^{y_n}\right)$

$$= \frac{1}{2^n} \sum_{\vec{y} \in \{0,1\}^n} \underbrace{(-1)^{x_1 y_1} (-1)^{x_2 y_2} \cdots (-1)^{x_n y_n}}_{\substack{= (-1)^{x_1 y_1 + x_2 y_2 + \cdots + x_n y_n} \\ = (-1)^{\vec{x} \cdot \vec{y}}}} \underbrace{Z^{y_1} \otimes Z^{y_2} \otimes \cdots \otimes Z^{y_n}}_{\hookrightarrow\ = Z^{\vec{y}}}$$

$$= \frac{1}{2^n} \sum_{\vec{y} \in \{0,1\}^n} (-1)^{\vec{x} \cdot \vec{y}} Z^{\vec{y}} \qquad \blacksquare$$

3. Purity of the output state : $\mathrm{Tr}[\rho^2]$ $\quad \rho = |\psi\rangle\langle\psi| \rightarrow \rho^2 = |\psi\rangle\overbrace{\langle\psi|\psi\rangle}^{1}\langle\psi| = \rho$

Theorem: States $\rho_1, \rho_2 \Rightarrow \mathrm{Tr}[\rho_1 \rho_2] = \mathrm{Tr}[F(\rho_1 \otimes \rho_2)]$ $\longleftrightarrow$ "Swap trick".

$$\boxed{F = \sum_{i,j=0}^{d-1} |j,i\rangle\langle i,j| \quad \rightarrow \text{swap operator.}}$$

$F|0,1\rangle = |1,0\rangle$

$F|2,3\rangle = |3,2\rangle$

Aside: swap operator.

- $F|i,j\rangle = |j,i\rangle$
- $F(|\psi\rangle \otimes |\phi\rangle) = F\left( \left( \sum_{i=0}^{d-1} \alpha_i |i\rangle \right) \otimes \left( \sum_{j=0}^{d-1} \beta_j |j\rangle \right) \right)$

$$= \sum_{i,j=0}^{d-1} \alpha_i \beta_j F(|i\rangle \otimes |j\rangle) \rightarrow |j\rangle \otimes |i\rangle$$

$$= \sum_{i,j=0}^{d-1} \alpha_i \beta_j |j\rangle \otimes |i\rangle$$

$$= \left( \sum_{j=0}^{d-1} \beta_j |j\rangle \right) \otimes \left( \sum_{i=0}^{d-1} \alpha_i |i\rangle \right)$$

$$= |\phi\rangle \otimes |\psi\rangle.$$

$F^\dagger = F \rightarrow$ check!

- For mixed states: $F(\rho \otimes \sigma) F = \sigma \otimes \rho$.

**Proof:** $\mathrm{Tr}[\,\cdot\,] = \sum_{k,l=0}^{d-1} \langle k,l|(\cdot)|k,l\rangle$

$$\Rightarrow \mathrm{Tr}[F(\rho_1\otimes\rho_2)] = \sum_{k,l=0}^{d-1}\langle k,l|F(\rho_1\otimes\rho_2)|k,l\rangle$$

$$= \sum_{k,l=0}^{d-1}\sum_{i,j=0}^{d-1} \underbrace{\langle k,l|i,j\rangle}_{=\langle k|i\rangle\langle l|j\rangle = \delta_{i,k}\delta_{l,j}} \underbrace{\langle j,i|\rho_1\otimes\rho_2|k,l\rangle}_{\langle j|\rho_1|k\rangle\langle i|\rho_2|l\rangle}$$

$$= \sum_{k,l=0}^{d-1} \langle l|\rho_1|k\rangle\langle k|\rho_2|l\rangle$$

$$= \sum_{l=0}^{d-1}\langle l|\rho_1\underbrace{\left(\sum_{k=0}^{d-1}|k\rangle\langle k|\right)}_{=\mathbb{1}}\rho_2|l\rangle$$

$$= \sum_{l=0}^{d-1}\langle l|\rho_1\rho_2|l\rangle$$

$$= \mathrm{Tr}[\rho_1\rho_2]. \quad \blacksquare$$

$|\vec{a}\rangle = |a_1, a_2, \dots, a_n\rangle$

$|\vec{b}\rangle = |b_1, b_2, \dots, b_n\rangle$

⊛ We can measure $\mathrm{Tr}[\rho^2]$ using Pauli expectation values!

For a system of $n$ qubits: $F = \sum_{\vec{a},\vec{b}\in\{0,1\}^n} |\vec{b},\vec{a}\rangle\langle\vec{a},\vec{b}|$

**Theorem:** $F = \frac{1}{2^n}\sum_{P\in\mathcal{P}_n} P\otimes P$ $\quad\left(\Rightarrow \mathrm{Tr}[\rho^2] = \frac{1}{2^n}\sum_{P\in\mathcal{P}_n}(\mathrm{Tr}[P\rho])^2\right)$

↳ $\mathrm{Tr}[F(\rho\otimes\rho)] = \frac{1}{2^n}\sum_P \mathrm{Tr}[(P\otimes P)(\rho\otimes\rho)]$

**Proof:** Recall that $\mathcal{P}_n = \{\mathbb{1}, X, Y, Z\}^{\otimes n}$ is a basis. So we must have

$$F = \frac{1}{4^n}\sum_{P_1,P_2\in\mathcal{P}_n} c_{P_1,P_2}\, P_1\otimes P_2, \text{ for some coefficients } c_{P_1,P_2}.$$

$$c_{P_1,P_2} = \langle P_1 \otimes P_2, F \rangle = \mathrm{Tr}\left[(P_1 \otimes P_2) F\right]$$ $\longrightarrow$ just apply the swap trick!

$$= \mathrm{Tr}\left[(P_1 \otimes P_2)\left(\sum_{\vec{x},\vec{y}\in\{0,1\}^n} |\vec{y},\vec{x}\rangle\langle\vec{x},\vec{y}|\right)\right]$$

$$= \sum_{\vec{x},\vec{y}\in\{0,1\}^n} \mathrm{Tr}\left[(P_1 \otimes P_2)\left(|\vec{y}\rangle\langle\vec{x}| \otimes |\vec{x}\rangle\langle\vec{y}|\right)\right]$$

$$= \sum_{\vec{x},\vec{y}\in\{0,1\}^n} \mathrm{Tr}\left[P_1 |\vec{y}\rangle\langle\vec{x}|\right] \mathrm{Tr}\left[P_2 |\vec{x}\rangle\langle\vec{y}|\right]$$

$$= \sum_{\vec{x},\vec{y}\in\{0,1\}^n} \langle\vec{x}|P_1|\vec{y}\rangle\langle\vec{y}|P_2|\vec{x}\rangle$$

$$= \sum_{\vec{x}\in\{0,1\}^n} \langle\vec{x}|P_1 \underbrace{\left(\sum_{\vec{y}\in\{0,1\}^n} |\vec{y}\rangle\langle\vec{y}|\right)}_{=\mathbb{1}} P_2|\vec{x}\rangle$$

$$= \mathrm{Tr}[P_1 P_2]$$

$$= \delta_{P_1,P_2} 2^n$$ $\longrightarrow$ b/c Paulis are orthogonal. ■

$$\hookrightarrow F = \frac{1}{4^n}\sum_{P_1,P_2} 2^n \delta_{P_1,P_2}\, P_1 \otimes P_2 = \sum_P \frac{1}{2^n} P \otimes P.$$

Suppose $P_1 = X \otimes Y \otimes Z \otimes \mathbb{1}$

$(n=4)$. $P_2 = Y \otimes Z \otimes Z \otimes X$

$\Rightarrow \mathrm{Tr}[P_1 P_2]$

$= \underbrace{\mathrm{Tr}[XY]}_{=0}\underbrace{\mathrm{Tr}[YZ]}_{=0}\underbrace{\mathrm{Tr}[Z^2]}_{=2}\underbrace{\mathrm{Tr}[X]}_{0}$

$= 0$

But $\mathrm{Tr}[P_1 \cdot P_1]$

$= \mathrm{Tr}[X^2]\mathrm{Tr}[Y^2]\mathrm{Tr}[Z^2]\mathrm{Tr}[\mathbb{1}]$

$= 2^4 \rightarrow 2^n.$

⊛ **Procedure for estimating Pauli expectation values:**

- Find spectral decomposition of $P$. $|\pm\rangle = \frac{1}{\sqrt{2}}(|0\rangle \pm |1\rangle)$ $|\pm i\rangle = \frac{1}{\sqrt{2}}(|0\rangle \pm i|1\rangle)$.

e.g., $P = X \otimes Y \otimes Z$

$\big( X = |+\rangle\langle+| - |-\rangle\langle-|,\ Y = |+i\rangle\langle+i| - |-i\rangle\langle-i|$

$Z = |0\rangle\langle0| - |1\rangle\langle1| \big)$

- Measure 1st qubit in X-basis
- Measure 2nd qubit in Y-basis
- Measure 3rd qubit in Z-basis.

$$= \left(|+\rangle\langle+| - |-\rangle\langle-|\right) \otimes \left(|+i\rangle\langle+i| - |-i\rangle\langle-i|\right) \otimes \left(|0\rangle\langle0| - |1\rangle\langle1|\right).$$

$$= \Big( |+\rangle\langle+| \otimes |+i\rangle\langle+i| - |+\rangle\langle+| \otimes |-i\rangle\langle-i| - |-\rangle\langle-| \otimes |+i\rangle\langle+i| + |-\rangle\langle-| \otimes |-i\rangle\langle-i| \Big) \otimes \left(|0\rangle\langle0| - |1\rangle\langle1|\right)$$

$$\begin{aligned}
= {} & |+\rangle\langle+| \otimes |+i\rangle\langle+i| \otimes |0\rangle\langle0| - |+\rangle\langle+| \otimes |+i\rangle\langle+i| \otimes |1\rangle\langle1| \\
& - |+\rangle\langle+| \otimes |-i\rangle\langle-i| \otimes |0\rangle\langle0| + |+\rangle\langle+| \otimes |-i\rangle\langle-i| \otimes |1\rangle\langle1| \\
& - |-\rangle\langle-| \otimes |+i\rangle\langle+i| \otimes |0\rangle\langle0| + |-\rangle\langle-| \otimes |+i\rangle\langle+i| \otimes |1\rangle\langle1| \\
& + |-\rangle\langle-| \otimes |-i\rangle\langle-i| \otimes |0\rangle\langle0| - |-\rangle\langle-| \otimes |-i\rangle\langle-i| \otimes |1\rangle\langle1|
\end{aligned}$$

Then the expectation value is

$$\begin{aligned}
\mathrm{Tr}[P\rho] = \mathrm{Tr}\Big[\Big( & |+\rangle\langle+| \otimes |+i\rangle\langle+i| \otimes |0\rangle\langle0| - |+\rangle\langle+| \otimes |+i\rangle\langle+i| \otimes |1\rangle\langle1| \\
& - |+\rangle\langle+| \otimes |-i\rangle\langle-i| \otimes |0\rangle\langle0| + |+\rangle\langle+| \otimes |-i\rangle\langle-i| \otimes |1\rangle\langle1| \\
& - |-\rangle\langle-| \otimes |+i\rangle\langle+i| \otimes |0\rangle\langle0| + |-\rangle\langle-| \otimes |+i\rangle\langle+i| \otimes |1\rangle\langle1| \\
& + |-\rangle\langle-| \otimes |-i\rangle\langle-i| \otimes |0\rangle\langle0| - |-\rangle\langle-| \otimes |-i\rangle\langle-i| \otimes |1\rangle\langle1| \Big)\rho\Big].
\end{aligned}$$

$$\begin{aligned}
= {} & \mathrm{Tr}\left[\left(|+\rangle\langle+| \otimes |+i\rangle\langle+i| \otimes |0\rangle\langle0|\right)\rho\right] \\
& - \mathrm{Tr}\left[\left(|+\rangle\langle+| \otimes |+i\rangle\langle+i| \otimes |1\rangle\langle1|\right)\rho\right] \\
& - \mathrm{Tr}\left[\left(|+\rangle\langle+| \otimes |-i\rangle\langle-i| \otimes |0\rangle\langle0|\right)\rho\right] \\
& + \mathrm{Tr}\left[\left(|+\rangle\langle+| \otimes |-i\rangle\langle-i| \otimes |1\rangle\langle1|\right)\rho\right] \\
& - \mathrm{Tr}\left[\left(|-\rangle\langle-| \otimes |+i\rangle\langle+i| \otimes |0\rangle\langle0|\right)\rho\right] \\
& + \mathrm{Tr}\left[\left(|-\rangle\langle-| \otimes |+i\rangle\langle+i| \otimes |1\rangle\langle1|\right)\rho\right] \\
& + \mathrm{Tr}\left[\left(|-\rangle\langle-| \otimes |-i\rangle\langle-i| \otimes |0\rangle\langle0|\right)\rho\right] \\
& - \mathrm{Tr}\left[\left(|-\rangle\langle-| \otimes |-i\rangle\langle-i| \otimes |1\rangle\langle1|\right)\rho\right].
\end{aligned}$$

⊛ Measurement outcome probabilities.

In general, $P = \sum_{\vec{x} \in \{0,1\}^n} \lambda_{\vec{x}} |v_{\vec{x}}\rangle\langle v_{\vec{x}}|$ , $\underline{\underline{\lambda_{\vec{x}} \in \{+1,-1\}}}$.

$$\Rightarrow Tr[P\rho] = \sum_{\vec{x} \in \{0,1\}^n} \underbrace{\lambda_{\vec{x}}}_{\text{Eigenvalues (known).}} \underbrace{Tr[|v_{\vec{x}}\rangle\langle v_{\vec{x}}|\rho]}_{\text{probabilities } p_{\vec{x}}}$$

probabilities $p_{\vec{x}}$ — This is what we observe!

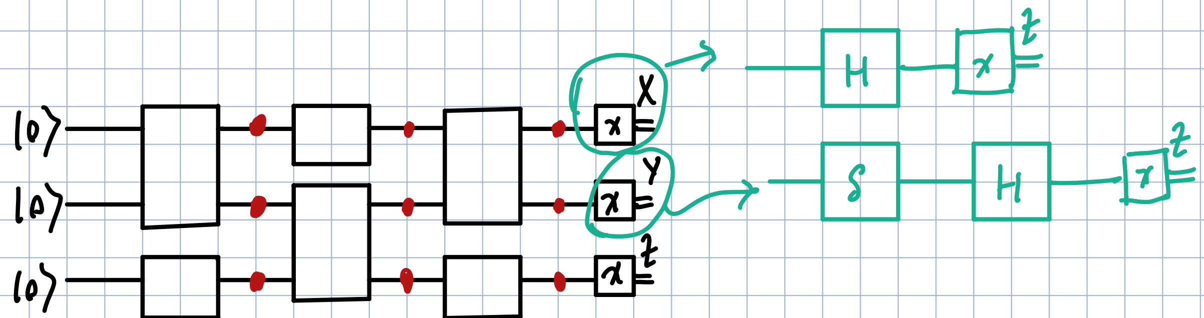
- Write X- and Y-basis measurements as Z-basis measurements.

$|\pm\rangle = H|0/1\rangle \quad \rightarrow \quad H = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$ (Hadamard gate).

$\Rightarrow H|\pm\rangle = |0/1\rangle \Rightarrow Tr[\rho |+\rangle\langle +|] = Tr[H\rho H |0\rangle\langle 0|]$.

$|\pm i\rangle = SH|0/1\rangle \longrightarrow S = \begin{pmatrix} 1 & 0 \\ 0 & i \end{pmatrix}$ (Phase gate)

$\Rightarrow HS|\pm i\rangle = |0/1\rangle \rightarrow Tr[\rho |+i\rangle\langle +i|] = Tr[HS\rho SH |0\rangle\langle 0|]$




- Estimator: Measurement outcome is a bit-string $\vec{x} \in \{0,1\}^n$.

    - Eigenvalues $\lambda_{\vec{x}}$ are known

    - Let $\hat{X}(\vec{x}) = \lambda_{\vec{x}}$ be the estimator (i.e., just output the corresponding eigenvalue!)

$X$ RV
$\mathbb{E}[X] = \mu$

⊛ $\hat{X}$ is a random variable with $Pr[\hat{X} = \lambda_{\vec{x}}] = Tr[|v_{\vec{x}}\rangle\langle v_{\vec{x}}|\rho]$.

⊛ $\hat{X}$ is an unbiased estimator of $\mathrm{Tr}[P\rho]$:

$\mathbb{E}[X] = \sum_x p(x) x$

$$\mathbb{E}[\hat{X}] = \sum_{\vec{x}} \underbrace{\Pr[\hat{X}=\lambda_{\vec{x}}]}_{\mathrm{Tr}[\rho|v_{\vec{x}}\rangle\langle v_{\vec{x}}|]} \lambda_{\vec{x}} = \sum_{\vec{x}\in\{0,1\}^n} \mathrm{Tr}[\rho|v_{\vec{x}}\rangle\langle v_{\vec{x}}|]\cdot\lambda_{\vec{x}} = \mathrm{Tr}\Big[\rho\Big(\underbrace{\sum_{\vec{x}\in\{0,1\}^n}\lambda_{\vec{x}}|v_{\vec{x}}\rangle\langle v_{\vec{x}}|}_{P}\Big)\Big]$$

$$= \mathrm{Tr}[\rho P].$$

- Algorithm:

For $i = 1, 2, \ldots, N$: ($N \equiv$ # of samples).

1. Prepare the state $\rho$
2. Measure each qubit in the basis defined by $P$.
3. Collect the outcome $\vec{x}_i$

Take the sample mean: $\hat{X}_N^{avg} := \frac{1}{N}\sum_{i=1}^{N}\hat{X}(\vec{x}_i) = \frac{1}{N}\sum_{i=1}^{N}\lambda_{\vec{x}_i}$

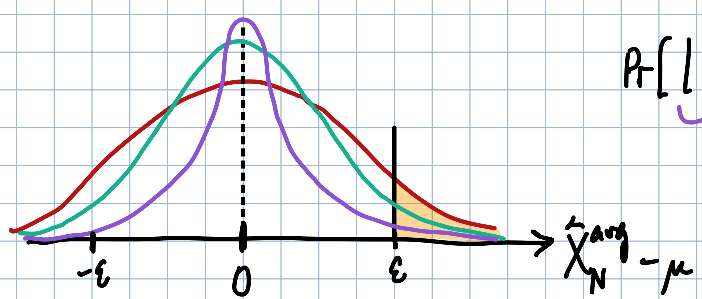
- Algorithm analysis: How good is the estimate? Use the law of large numbers.

1. Theorem (Law of large numbers) (informal)

   Sample mean converges to true mean as $N\to\infty$.

2. The true mean is $\underbrace{\mathbb{E}[\hat{X}]}_{\equiv\mu} = \mathrm{Tr}[P\rho]$ (b/c $\hat{X}$ is an unbiased estimator).

3. Convergence rate: Consider the "tail probability"




$\Pr[\underbrace{|\hat{X}_N^{avg} - \mu|}_{\text{estimate error.}} \geq \varepsilon]$. } Probability that the error is high.

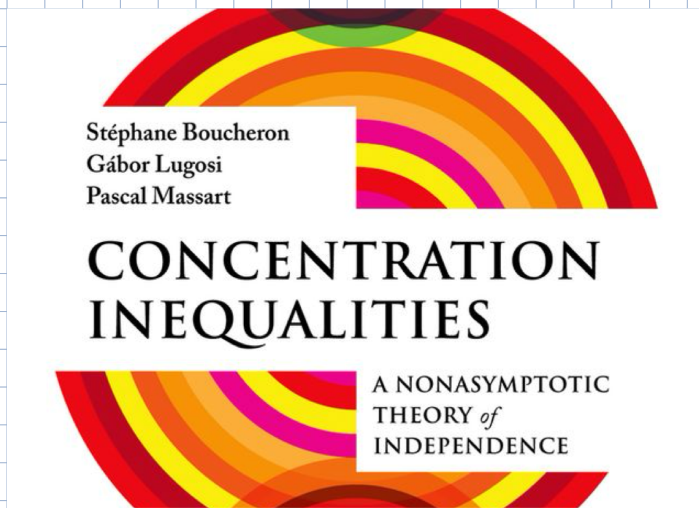
Apply the Hoeffding inequality:

**Theorem 2.8** (HOEFFDING'S INEQUALITY) *Let $X_1, \ldots, X_n$ be independent random variables such that $X_i$ takes its values in $[a_i, b_i]$ almost surely for all $i \le n$. Let*

$$S = \sum_{i=1}^{n} (X_i - \mathbf{E}X_i) .$$

*Then for every $t > 0$,*

$$\mathbf{P}\{|S| \ge t\} \le 2\exp\left(-\frac{2t^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right) .$$




(*) Here, we have $X_i \equiv \frac{1}{n}\hat{X}$, $a_i = -\frac{1}{n}$, $b_i = \frac{1}{n}$

$$\Rightarrow \Pr\left[\,|\hat{X}_N^{avg} - \mu| \geq \varepsilon\right] \leq 2e^{-\frac{N\varepsilon^2}{2}}$$

$$\Rightarrow \Pr\left[\,|\hat{X}_N^{avg} - \mu| \leq \varepsilon\right] = 1 - \Pr\left[\,|\hat{X}_N^{avg} - \mu| \geq \varepsilon\right]$$

$$\geq 1 - 2e^{-N\varepsilon^2/2}$$

$$\to 1 \text{ as } N \to \infty .$$

(*) Exponential convergence at rate $\varepsilon^2/2$.

4. How many samples do we need to get $\varepsilon$ error w/ prob. at least $1-\delta$?

$$\Pr\left[\left|\hat{X}_N^{avg} - \mu\right| \le \varepsilon\right] \geq 1 - 2e^{-N\varepsilon^2/2} \geq 1-\delta$$

$$\Rightarrow \delta \geq 2e^{-N\varepsilon^2/2} \geq e^{-N\varepsilon^2/2}$$

$$\Rightarrow \log(\delta) \geq -N\left(\varepsilon^2/2\right)$$

$$\Rightarrow N \geq \frac{-\log(\delta)}{\varepsilon^2/2}$$

$$\Rightarrow N \gtrsim \frac{\log(1/\delta)}{\varepsilon^2}$$